Services for Distributed Online Data Repositories

Sara J. Graves, Ph.D.

Director, Information Technology and Systems Center
University Professor, Computer Science Department
University of Alabama in Huntsville
Director, Information Technology Research Center
National Space Science and Technology Center
256-824-6064
sgraves@itsc.uah.edu

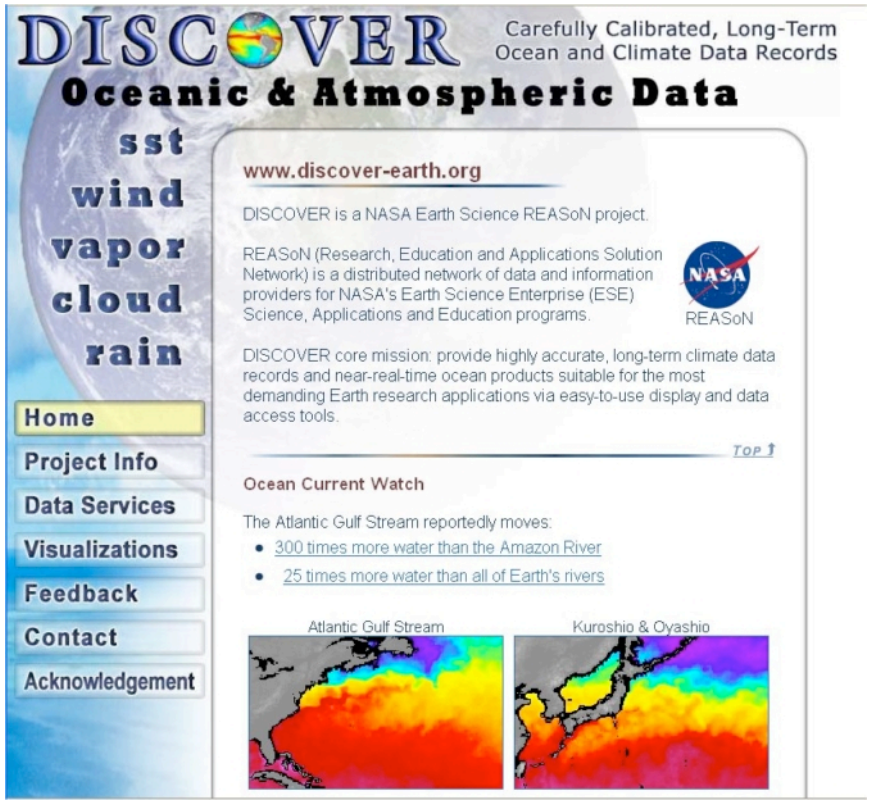
http://www.itsc.uah.edu

itsc

UAH
The University of Alabama in Huntsville

Distributed Information Services: Climate/Ocean Products for Earth Research (DISCOVER)

- A REASoN collaboration between Remote Sensing Systems, NASA/MSFC and the University of Alabama in Huntsville
  - 5 year project starting April 2004
- Science goals
  - Multi satellite inter-calibration and cross-validation
  - Highly accurate, long term ocean and climate products
- IT goals
  - On-line services for data access and visualization
  - Interoperability technologies for improved data usability
  - Flexible architecture to adapt to changing user requirements



http://www.discover-earth.org/

itsc

UAH
The University of Alabama in Huntsville

Key Personnel

- Project Coordinator
  - Michael Goodman – Co-I, NASA/MSFC
- Atmospheric Science
  - Frank Wentz – PI, RSS
  - Roy Spencer – Co-I, UAH
- Information Technology
  - Sara Graves – Co-I, UAH
  - Helen Conover – technical lead
  - Ken Keiser – data pool development

http://datapool.nsstc.nasa.gov

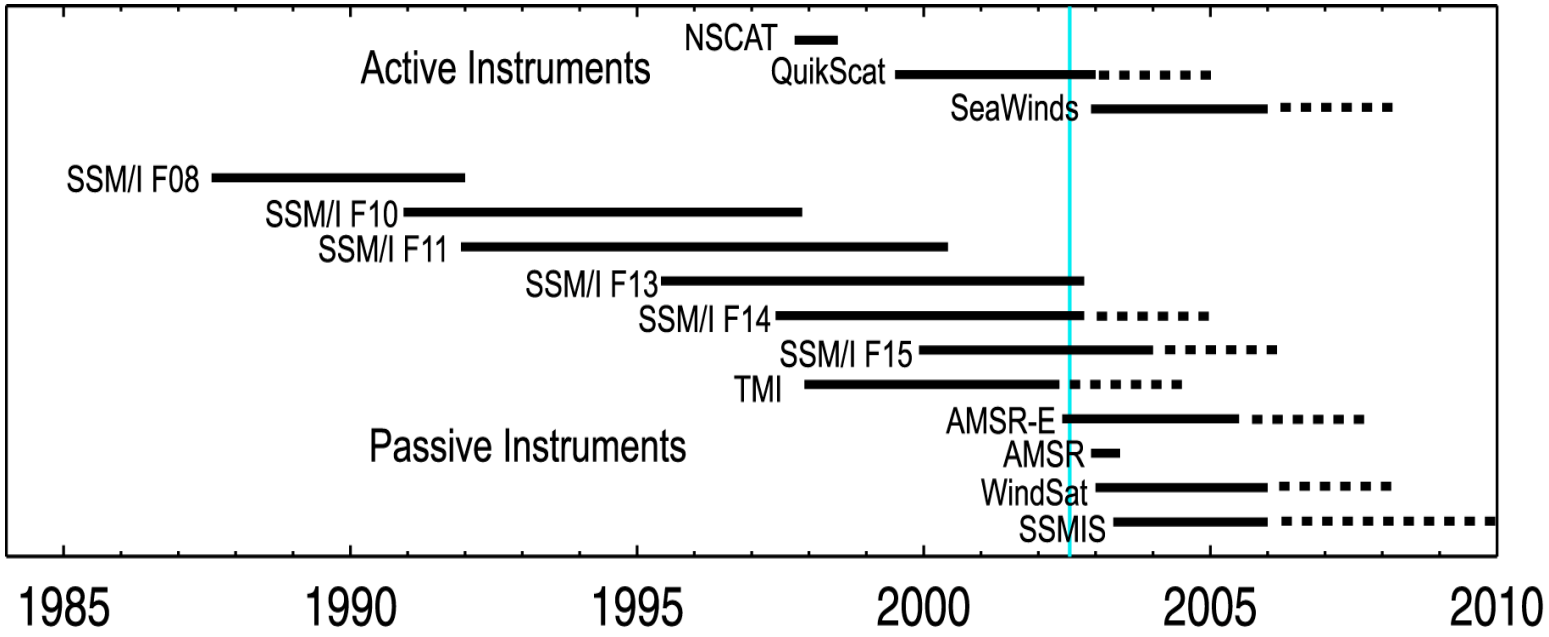
itsc

UAH
The University of Alabama in Huntsville

Long Term Systematic Measurements

Merging multiple sensors from multiple platforms into geophysical data sets consistent in both space and time to produce highly accurate, long-term ocean and climate products

- 27 microwave satellite instruments from the past, present, and future

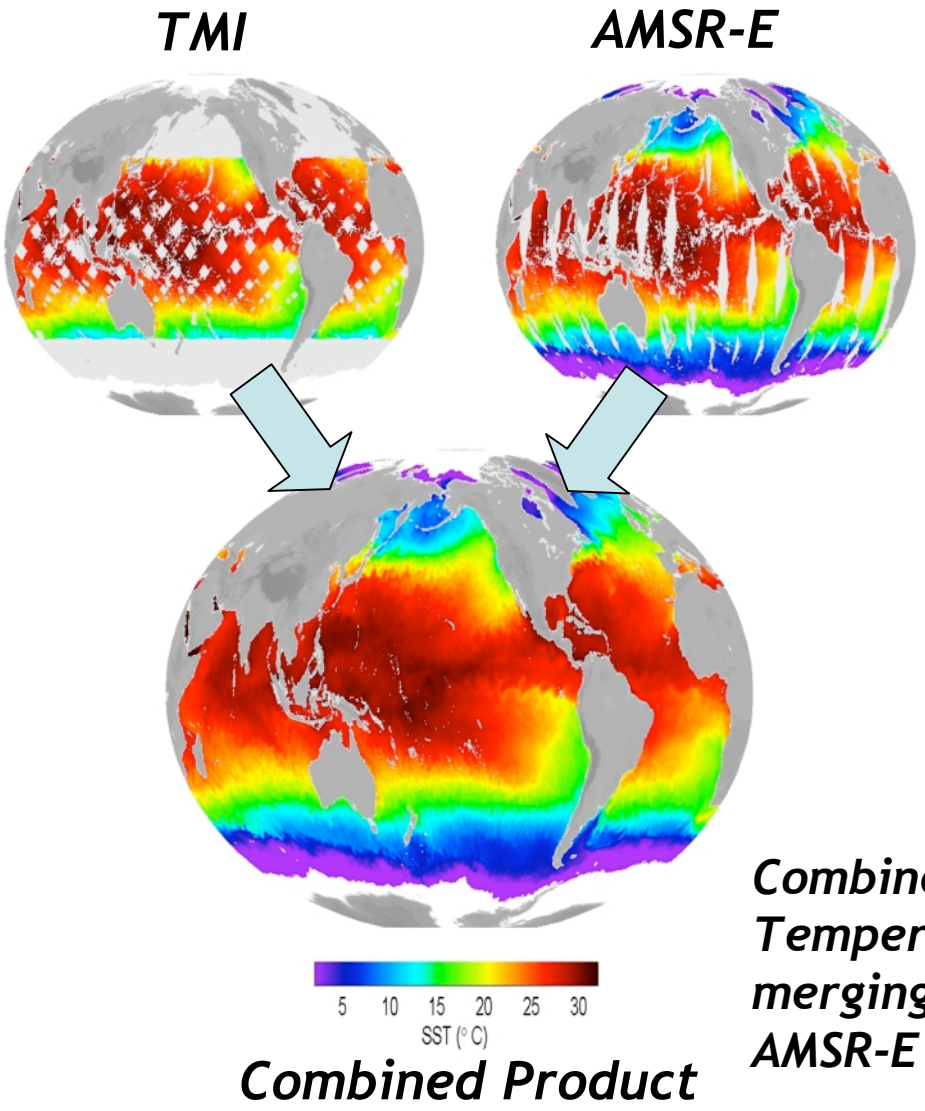


Periods of operation for microwave instruments that can retrieve ocean surface winds

itsc

UAH
The University of Alabama in Huntsville

DISCOVER Data Products



Combined Product

Data products include:

- Brightness temperatures
- Sea surface temperature
- Wind speed
- Air temperature
- Atmospheric water vapor
- Cloud liquid water
- Rain rate

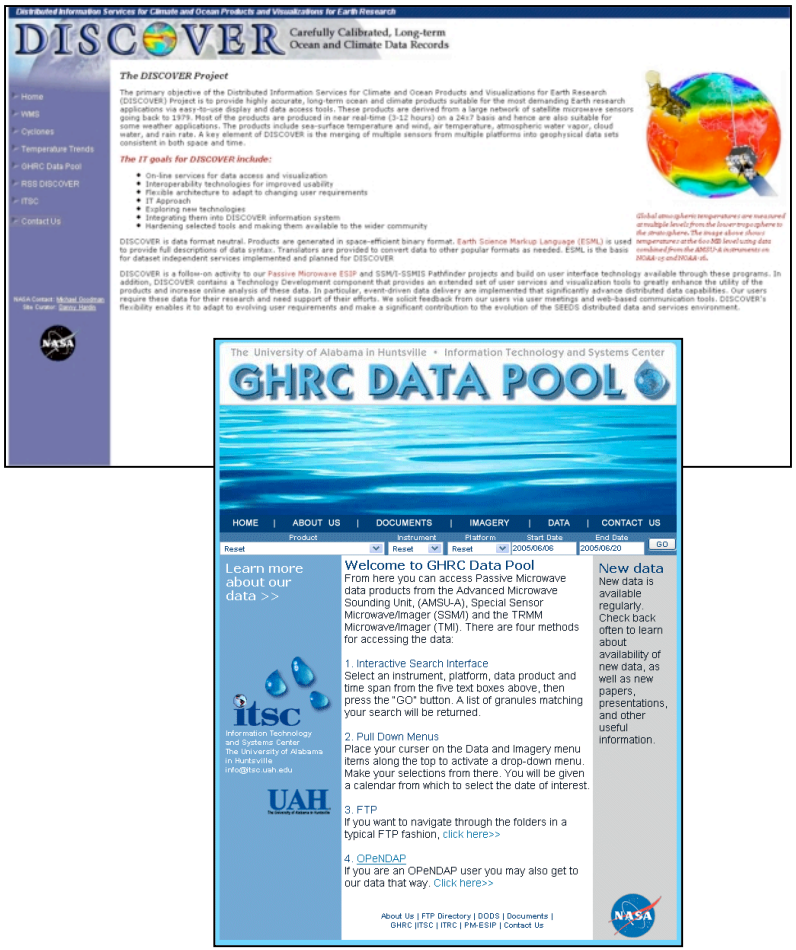
Combined Sea Surface Temperature product merging data from TMI and AMSR-E available now

itsc

UAH
The University of Alabama in Huntsville

IT for DISCOVER

- IT goals
  - On-line services for data access and visualization
  - Interoperability technologies for improved usability
  - Flexible architecture to adapt to changing user requirements
- IT Approach
  - Exploring new technologies
  - Integrating them into DISCOVER information system
  - Hardening selected tools and making them available to the wider community

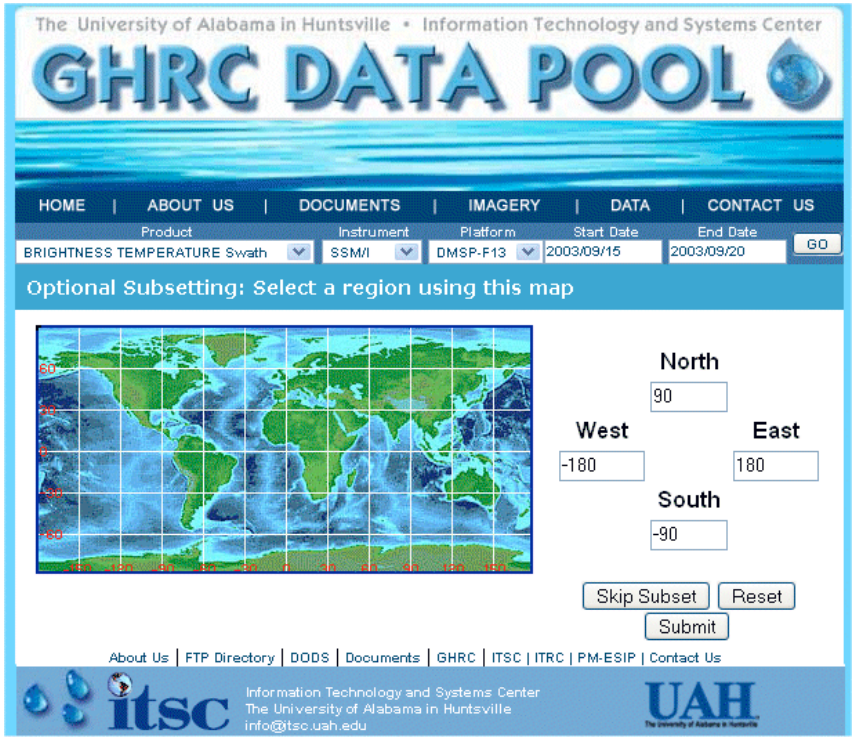


itsc

UAH
The University of Alabama in Huntsville

Data Management Objectives

- Distributed Service Architecture
  - Custom Order Processing
  - Modular Software Components
  - Service Semantics
- Open, Distributed, Heterogeneous Data System
  - Improved Access and Use of Data and Information Products
  - Extends Online Data Distribution



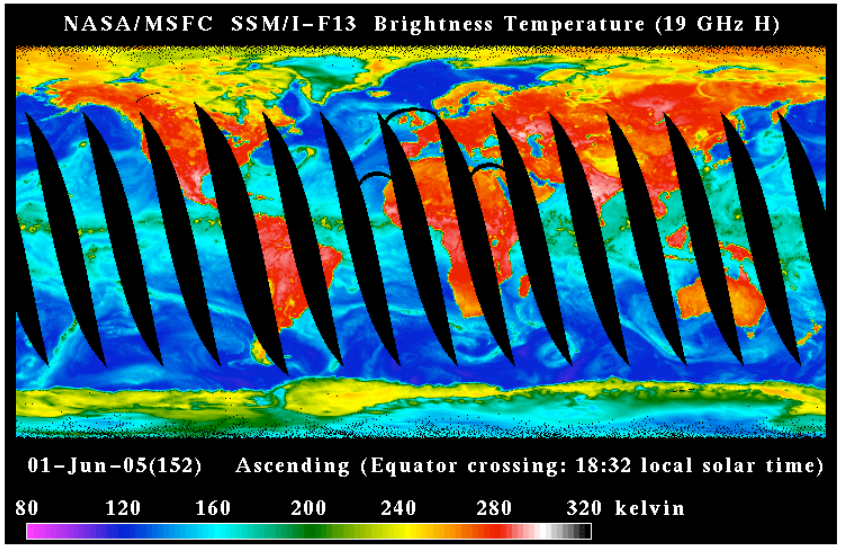
http://datapool.nsstc.nasa.gov

itsc

UAH
The University of Alabama in Huntsville

Data Pool Services Provide

- Improved access to online repositories
- Automated management and distribution
- Support of multiple interfaces
  - HTTP/FTP
  - OPeNDAP/DODS
  - OGC Web Services
  - Manual requests
- Automated electronic retrieval of data
- Automated order tracking/metrics
- Implementation across heterogeneous platforms

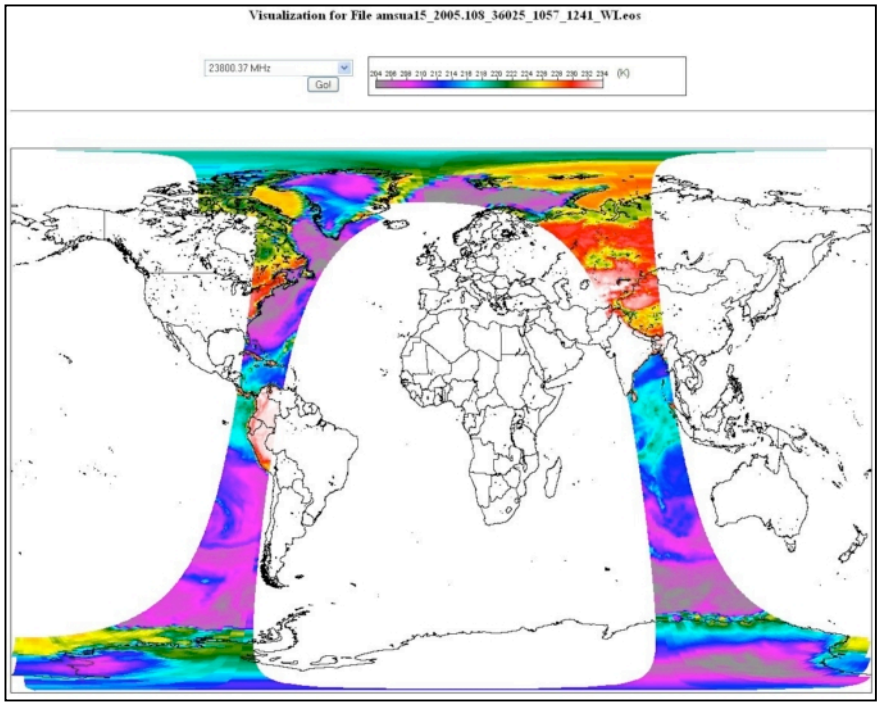


itsc

UAH
The University of Alabama in Huntsville

GHRC Data Pool Features

- Data Ingest through catalog services
- Search and Order Interfaces
- Order Tracking services
- Online Repository
- Subsetting
- Visualization through web mapping services
- Order broker
- Packaging services

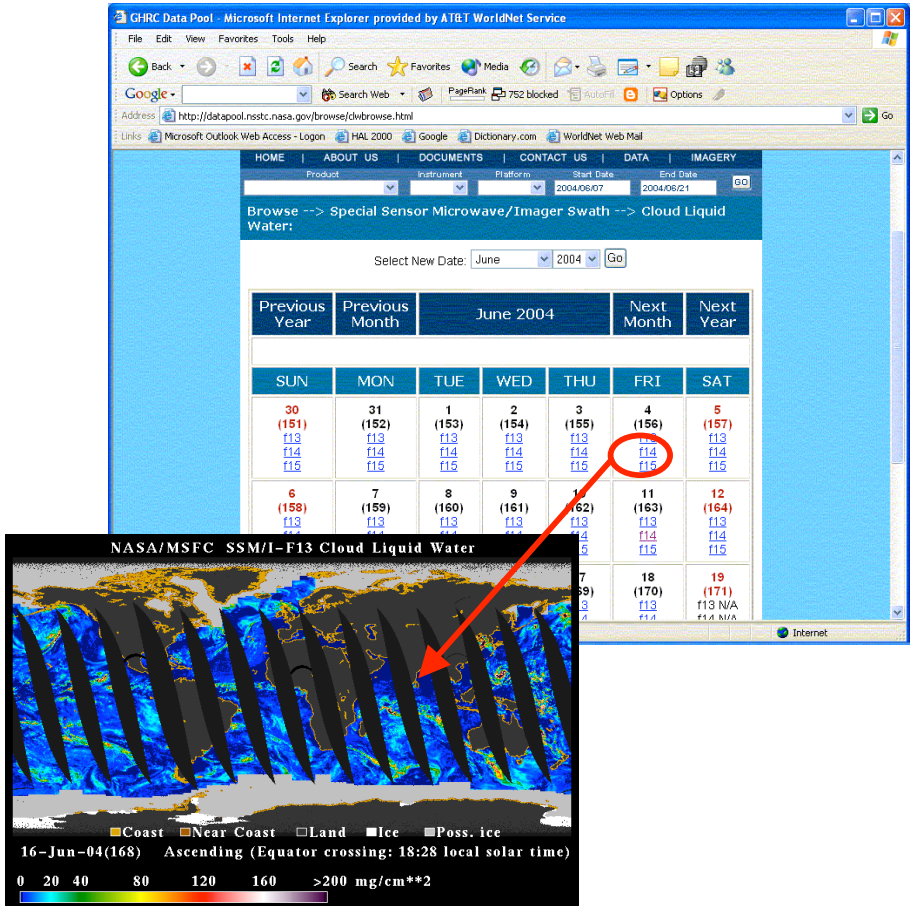


itsc

UAH
The University of Alabama in Huntsville

GHRC Data Pool Features

- On-line data access with integrated data services
- Automated ordering, visualization, packaging and delivery of scientific data
- Multiple distributed repositories at UAH and RSS
- Common user interface, data catalog and order tracking

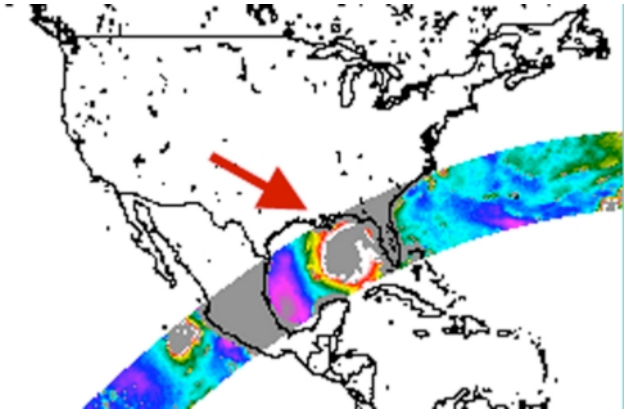


itsc

UAH
The University of Alabama in Huntsville

GHRC Data Pool Features

- Data Pool Navigation Options:
  - Search via keyword and geotemporal location
  - Browse via calendar
  - Browse data directories
  - Provides Unidata THREDDS catalog

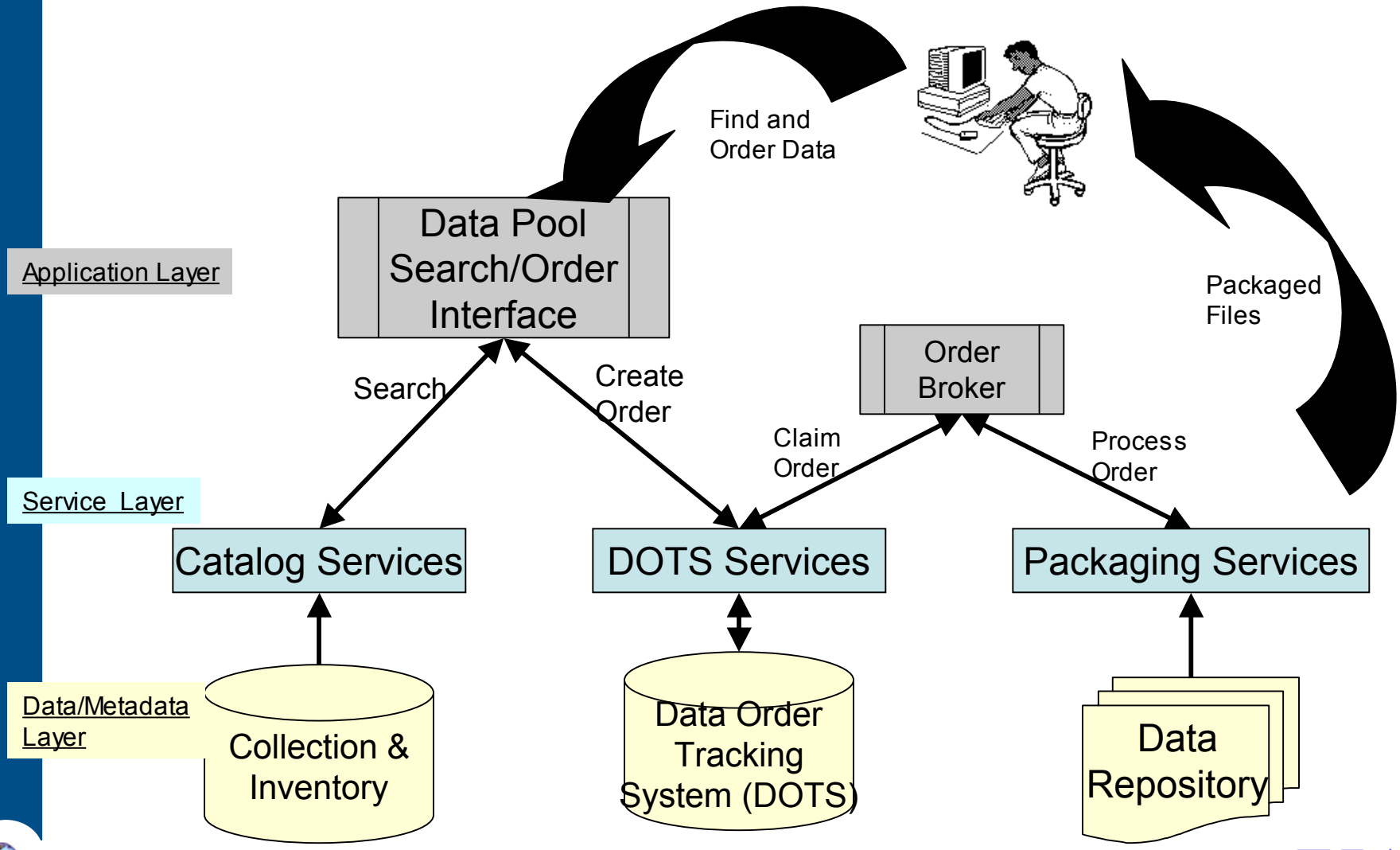


- Data Access Services:
  - FTP and HTTP
  - OPeNDAP
  - OGC WMS and WCS
- Data Packaging Services:
  - Subsetting
  - Re-formatting
  - Bundling

itsc

UAH
The University of Alabama in Huntsville

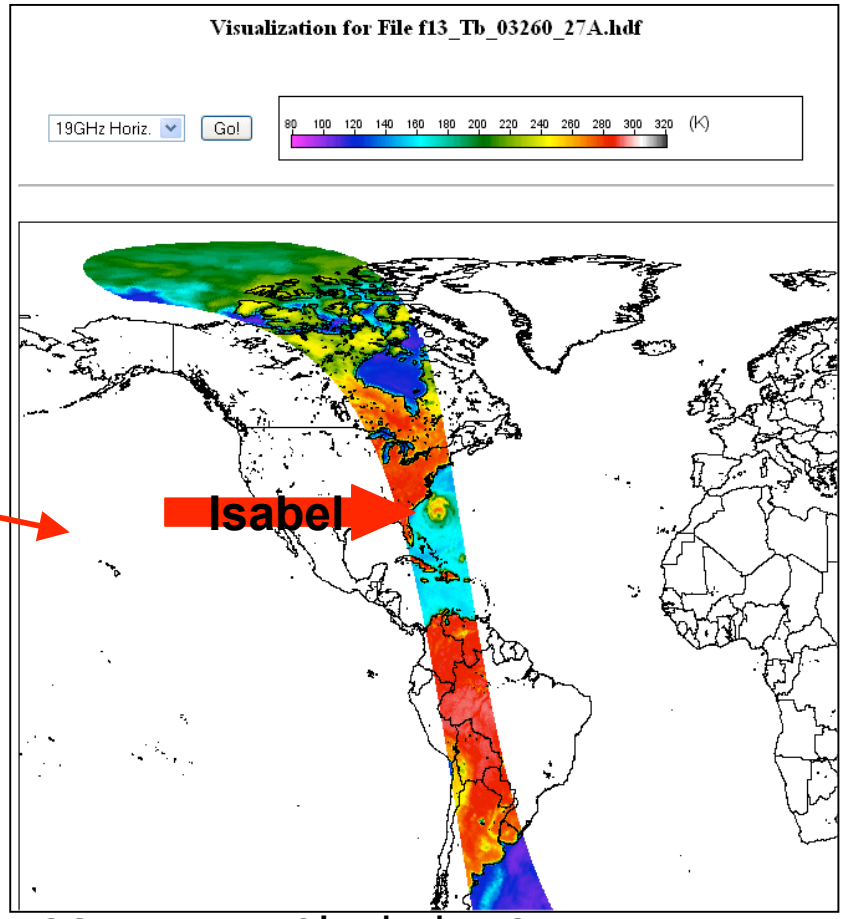
Data Pool Services Interactions



itsc

UAH
The University of Alabama in Huntsville

Integration of OGC Web Services



SSM/I View of Isabel in September 2003

Interactive, real-time browse images through use of OGC web mapping services

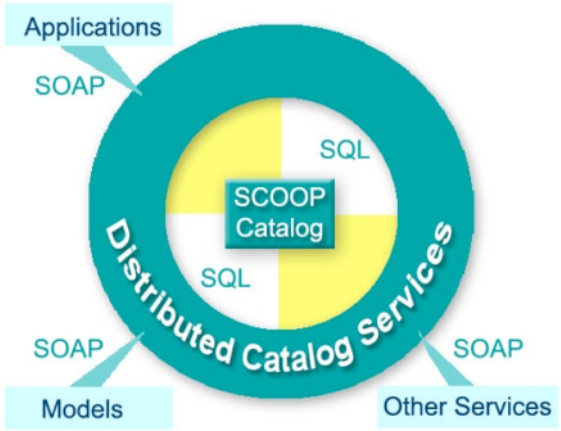
itsc

UAH
The University of Alabama in Huntsville

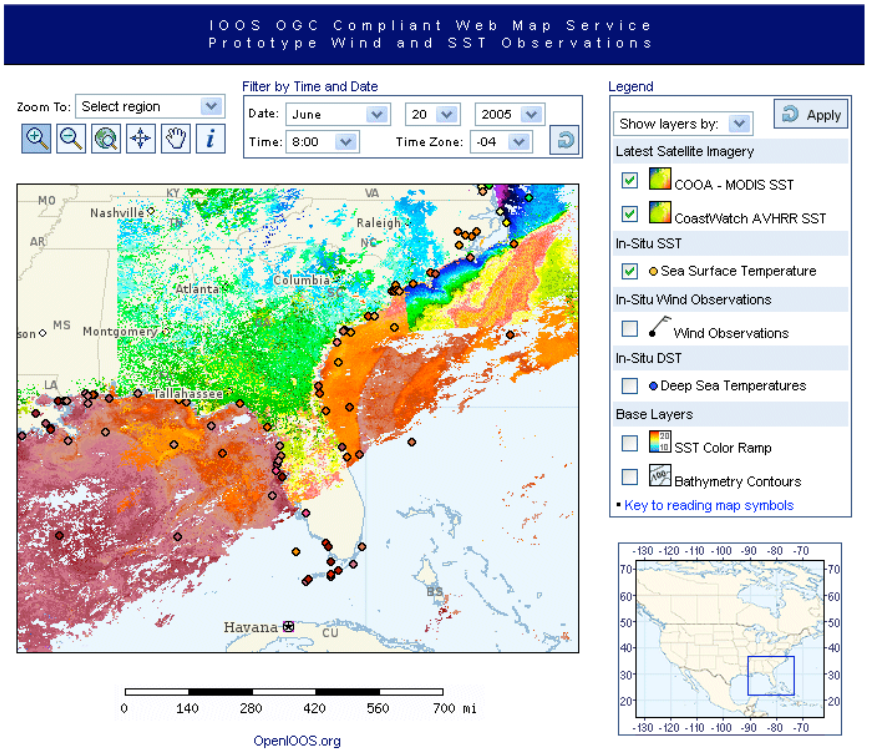
Technology Transfer Example

The SURA Coastal Ocean Observing and Prediction (SCOOP) program is incorporating many of the Data Pool approaches in the SCOOP Data System. Main enhancements include:

- Data Transport to users and archives via Unidata's Local Data Manager (LDM)
- Catalog services for all data management functions



http://www.itsc.uah.edu/scoop



http://www.openioos.org

itsc

UAH
The University of Alabama in Huntsville

DISCOVER Information Products: Mining to Detect Tropical Cyclones

Advanced Microwave Sounding Unit (AMSU-A) Data

Calibration/ Limb Correction/ Converted to Tb

Data Archive

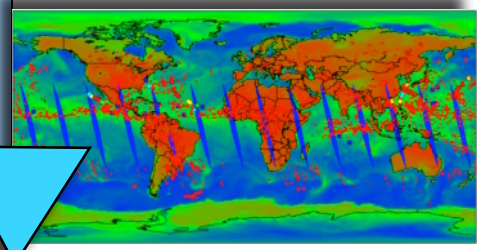
ADaM Mining Environment

Mining Plan:

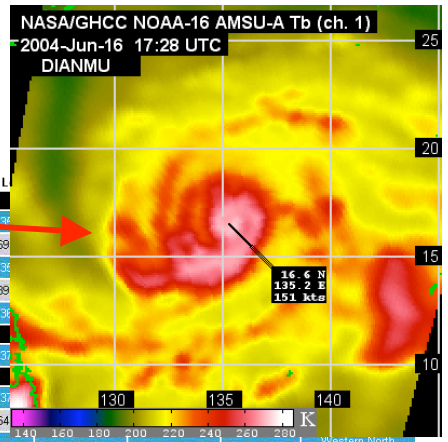
- Water cover mask to eliminate land
- Laplacian filter to compute temperature gradients
- Science Algorithm to estimate wind speed
- Contiguous regions with wind speeds above a desired threshold identified
- Additional test to eliminate false positives
- Maximum wind speed and location produced

Knowledge Base

Further Analysis



Result



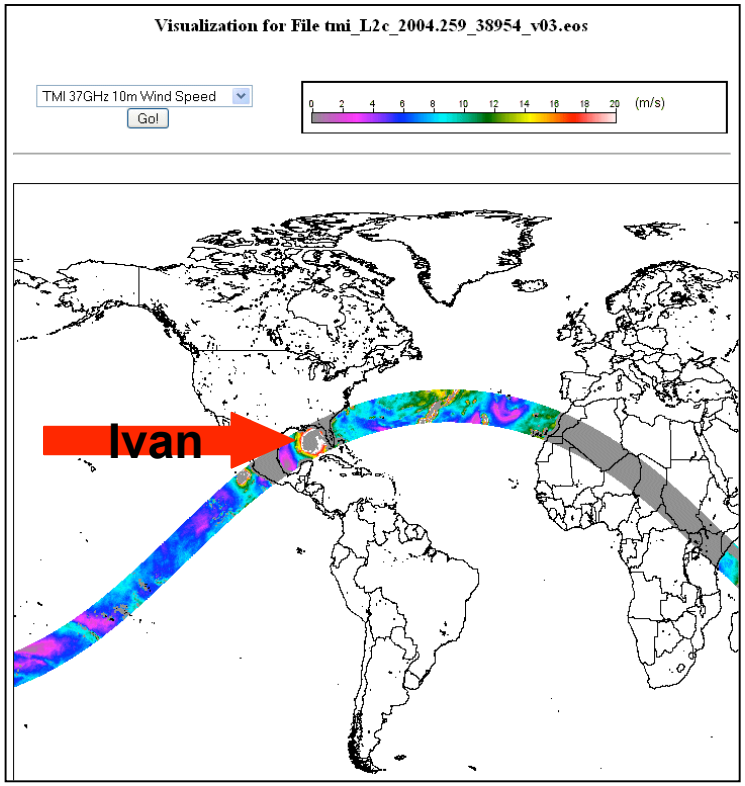
Results are placed on the web, made available to National Hurricane Center & Joint Typhoon Warning Center, and stored for further analysis

itsc

UAH
The University of Alabama in Huntsville

Conclusions

- Use of services to support distributed online repositories is effective.
- Distributed services insulate customers and applications from underlying computational systems
- Use of distributed services protect applications from ripples of underlying service changes
- Service use promotes the ability to scale systems across multiple repositories



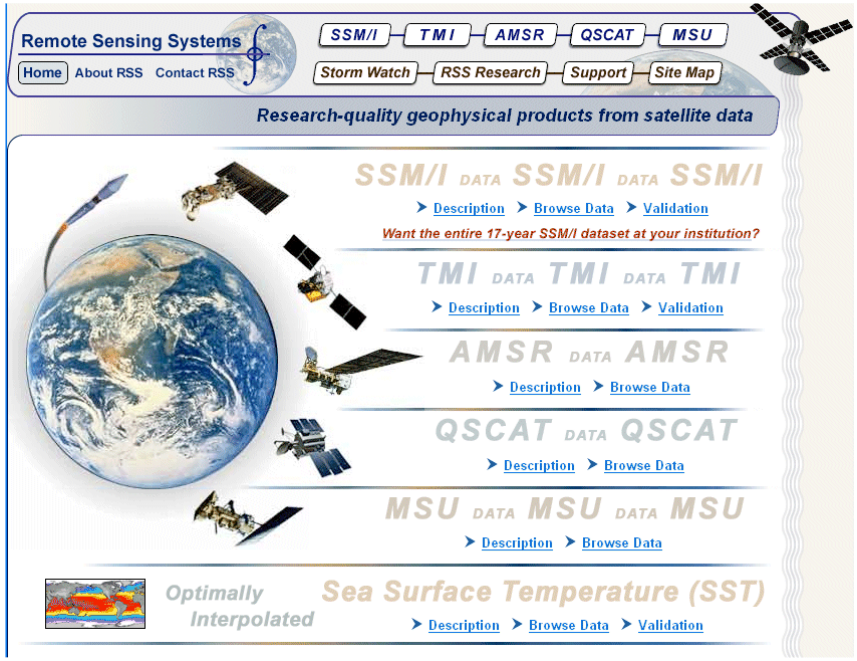
TMI View of Ivan in September, 2004

itsc

UAH
The University of Alabama in Huntsville

Next Steps

- Incorporate additional repositories
  - RSS datasets into GHRC catalog
- Provide additional services
  - OPeNDAP/ESML
  - WMS at RSS
- Provide GCMD entries for RSS datasets
- Investigate adaptation to other domains



http://www.remss.com

itsc

UAH
The University of Alabama in Huntsville

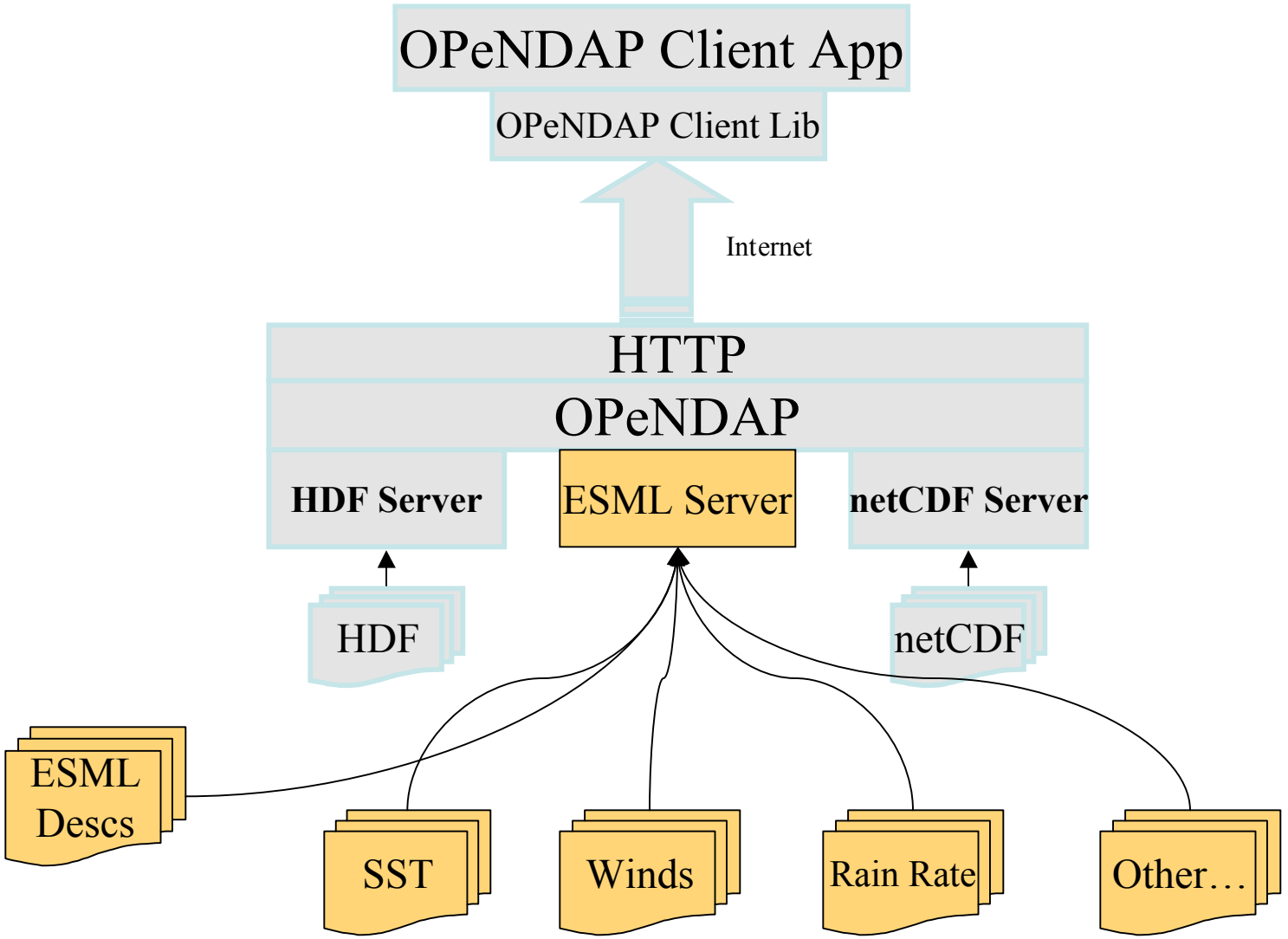
OPeNDAP-ESML Data Server

- OPeNDAP data transport
  - Servers for each data type
  - Clients for each application
- ESML adds layer of abstraction to data server
  - Single ESML-OPeNDAP server for multiple data types
  - Simplifies set-up and maintenance for data provider
  - New OPeNDAP server not required if ESML already supports a data type that OPeNDAP doesn't – e.g., GeoTIFF
- Leveraging two maturing, NASA-funded technologies to provide added interoperability

itsc

UAH
The University of Alabama in Huntsville

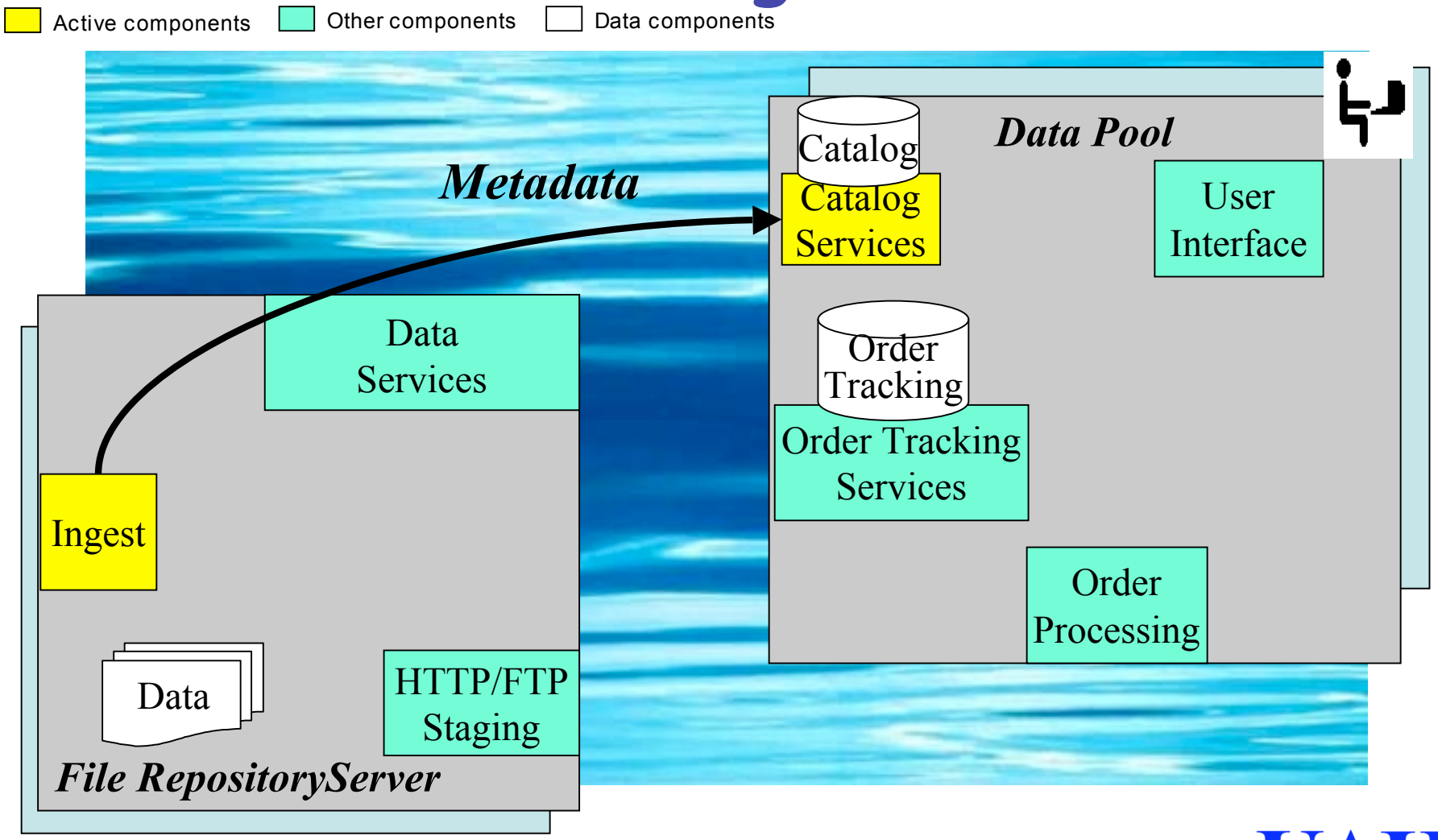
ESML-OPeNDAP Server



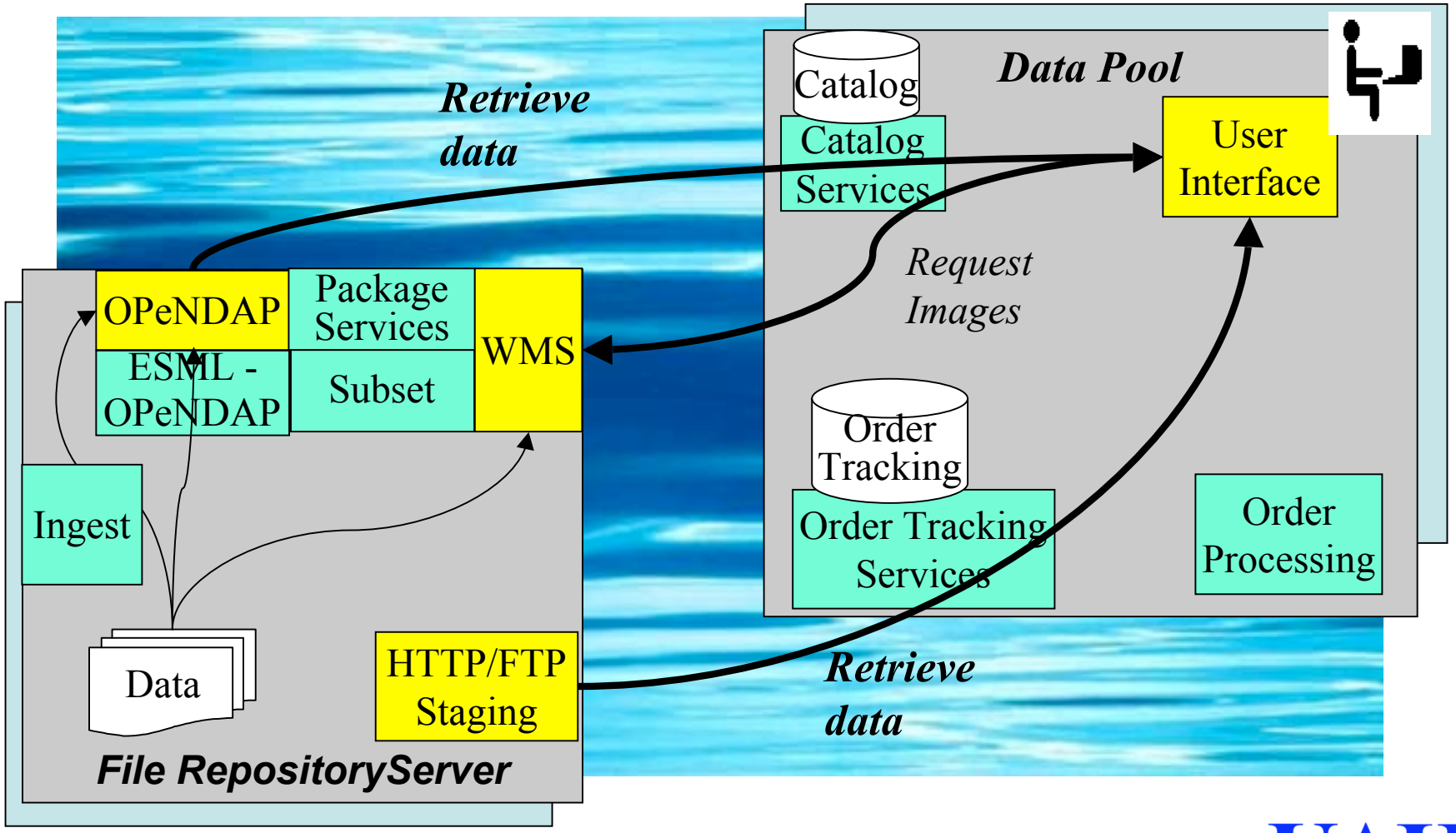
itsc

UAH
The University of Alabama in Huntsville

Data Pool – Architecture
Data Ingest



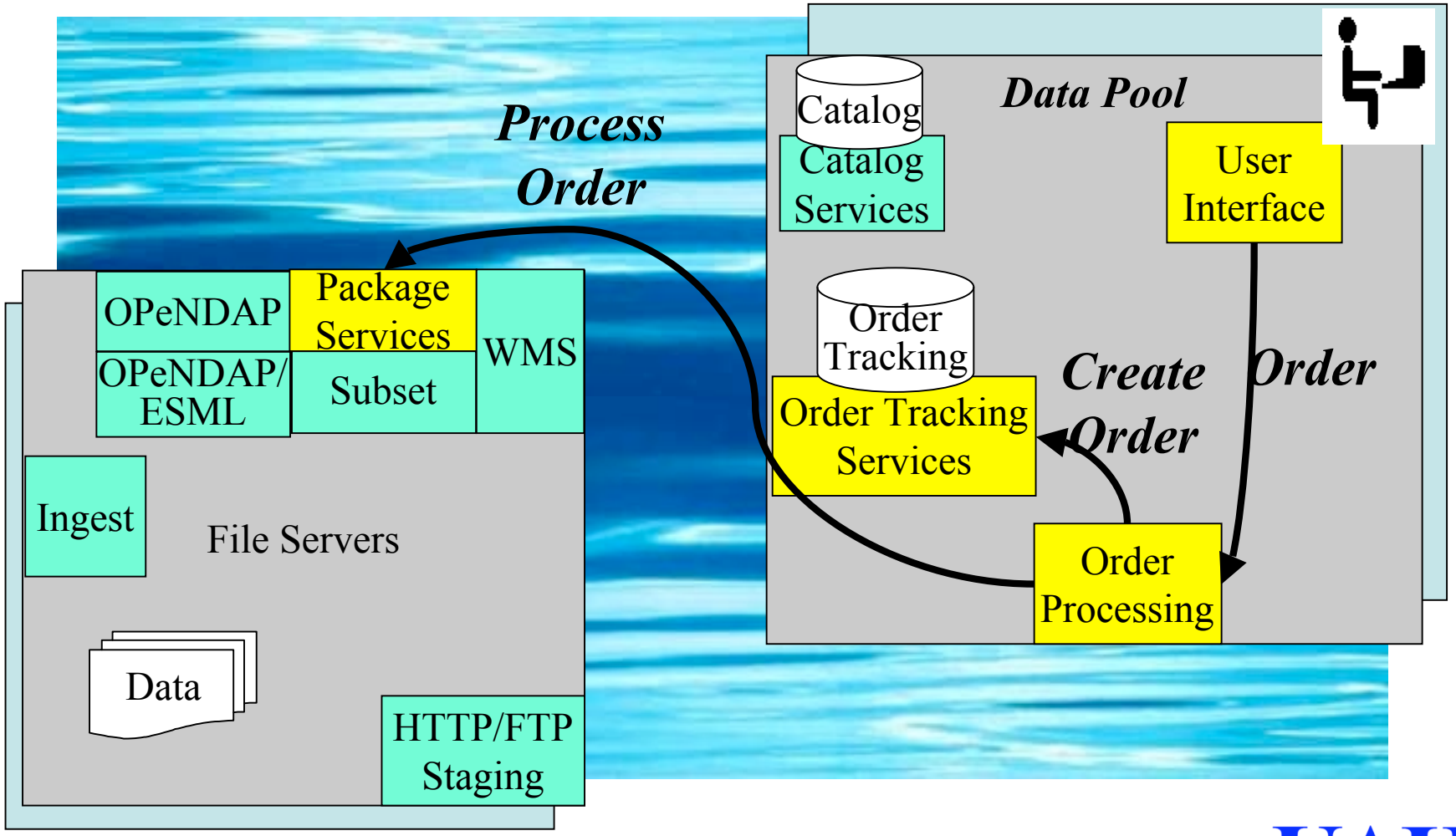
Data Pool – Architecture
Data Access



itsc

UAH
The University of Alabama in Huntsville

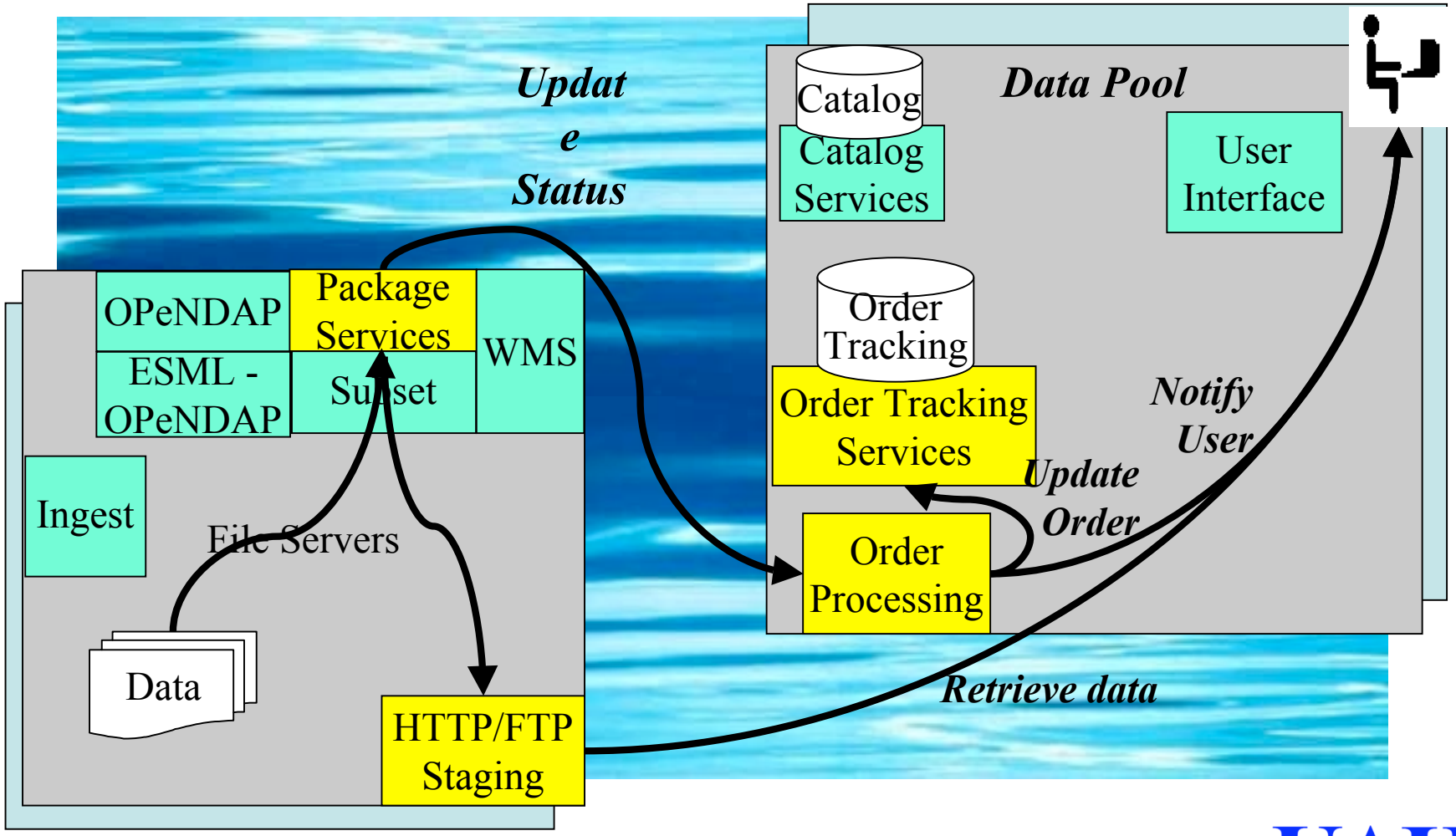
Data Pool – Architecture Generating Order Request



itsc

UAH
The University of Alabama in Huntsville

Data Pool – Architecture Processing Order



itsc

UAH
The University of Alabama in Huntsville